

## Tsar Reference Guide

This Reference Guide describes the theory and methods behind the following techniques used in Tsar:

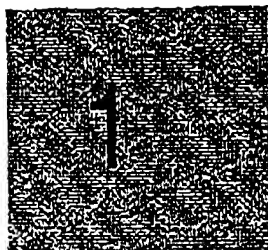
- property calculations
- Tsar statistics
- data reduction techniques
- clustering techniques
- regression techniques.

17.Déc. 2002 12:11

020 7240 7258

No. 9861

P. 3/24



## Property Calculations

### Overview

The property calculations in Tsar derive a wide range of structural descriptors from the simple 2D and 3D structural information available for a structure. In addition, some calculations use atomic partial charges to derive further properties.

- ❖ Refer to Chapter 8, "*Calculating Structure Properties*" in the Tsar User Guide, for information about using Tsar to calculate structure properties.

You can also import structure information directly from databases. Tsar is supplied with two standard databases of substituent properties taken from the published work of Hansch and Leo (*aromatic.tsard* and *aliphatic.tsard*). A further database provided by Dr. van de Waterbeemd is stored as *waterbeemd.tsard*.

- ❖ Refer to Chapter 5, "*Managing Data*" in the Tsar User Guide, for information about using Tsar to create your own databases and import information from databases.

### Contents

Structure property calculations	1-2	Kier and Hall shape indices	1-18
Moments of inertia	1-4	Counting H-bond donors and acceptors	1-23
Molar refractivity	1-4	Donor atoms	1-23
log P	1-5	Acceptor atoms	1-23
Verloop parameters	1-6	Autocorrelograms	1-24
Dipole moments	1-7	2D autocorrelogram	1-24
Charge-2	1-9	3D autocorrelogram	1-24
Topological, connectivity, and shape indices	1-12	Weighting factors	1-25
Topological indices	1-12	Topliss trees	1-26
Connectivity indices	1-13	References	1-27
Kier and Hall electrotopological state indices	1-17		

## Chapter 1 - Property Calculations

# Structure property calculations

The structure property calculations are accessed and defined using options from the **Properties** menu. The calculations use the 2D and 3D structural information stored in the project to derive further property descriptors, such as:

- mass, surface area, volume
- Verloop parameters
- moments of inertia
- dipoles
- molar refractivity
- lipophilicity, lipoles
- connectivity, shape, topology, and electrotopology indices
- autocorrelograms
- similarity indices (Asp users only)
- electrostatic parameters (Vamp users only)
- toxicity calculations (*TOPKAT QSTRs* for Tsar users only).



Access online Help for Asp, Vamp, and *TOPKAT QSTRs* for Tsar by choosing **Help | Online Reference | Asp/Vamp/TOPKAT QSTRs for Tsar** from a Project Window.

The Verloop parameter and molecular volume calculations use approximate numerical algorithms, which may produce very slightly different results for the same molecule in different orientations.

The calculations make the following assumptions:

- molecular mass, moments of inertia/principal ellipsoid, and autocorrelogram calculations assume mean atomic masses for constituent atoms (parameterized for elements up to atomic number 100)

17.Déc. 2002 12:12

020 7240 7258

No. 9861

P. 5/24

Structure property calculations

- Verloop parameters, surface area, and volume calculations assume standard Van der Waals radii for each element
- surface area calculations use a probe radius of 1.4Å.

*Tsar Reference Guide*

I-3

*Topological, connectivity, and shape indices***Balaban Indices**

The Balaban index,  $J$ , is the average-distance sum connectivity (Balaban, 1982).

For a connected molecular graph  $G$ ,

$$J = \frac{M}{\mu + 1} \sum_{\text{all edges}} (D_i D_j)^{-0.5}$$

where:

$M$  = number of edges in  $G$ .

$\mu$  = cyclomatic number of  $G$ . On a polycyclic graph,  $\mu$  is the minimum number of edges that must be removed before  $G$  becomes acyclic.

$D_i = \sum_{j=1}^n D_{ij}$  and  $D_{ij}$  is as defined for the Wiener index.

**Connectivity indices**

Hall and Kier (1992) have developed molecular connectivity indices (Chi) that reflect the atom identities, bonding environments and number of bonding hydrogens. These Kier indices are consequently useful in a wider variety of applications (Hall and Kier, 1992 detail many examples).

Molecules that are drawn without hydrogen atoms can be decomposed into fragments of length  $m$ , which may be divided into different categories. Hall and Kier defined four series of fragment categories: Path, Cluster, Path/Cluster, and Ring. The spread and numbers of fragment membership for each category is determined by molecule connectivity.

## Chapter 1 - Property Calculations

### Electrotopological Indices calculated by Tsar

You can use the Topology options in Tsar to calculate the following indices:

**Sum of Atoms** calculate the sum of the states of all the individual atoms in the molecule or substituent (see diagram).

#### First Atom

calculate the state of the first substituent atom attached to the generic ipso atom

#### Ipsso Atom

calculate the state of the generic atom to which the chosen substituent is attached

#### Sum of Atoms (substituent)

calculate the sum of the states of all the individual atoms of the chosen substituent (substituent 2)

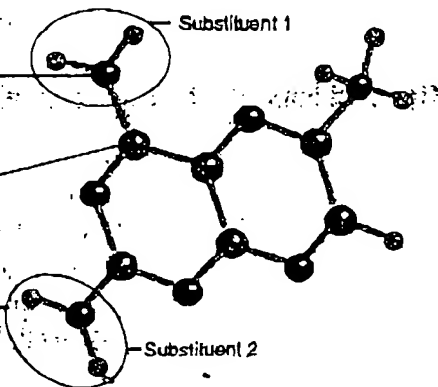


Diagram of a whole molecule with two substituents

### Kier and Hall shape indices

The Kappa index (Kier and Hall, 1992) is a molecule shape index based on the assumption that the shape of a molecule is a function of the number of atoms and their bonding relationship. This reference contains a detailed discussion of the applications of these indices, including a comparison with the Taft Steric parameter.

Notice that work was carried out on hydrogen suppressed or 'skeleton' structures.

Hall and Kier found that for molecules with  $A$  atoms, a set of path variables  ${}^mP$  could be defined:

${}^mP$  = the total number of paths along adjacent bonds of length  $m$  bond lengths, for each molecule.

They found that molecules with dissimilar bonding patterns had different values of  ${}^mP$  over the range of  $m$ . Also, that for each



17 Dec 2002 12:23

020 7240 7258

No. 9951

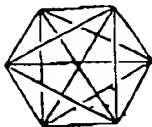


P. 21/24

Topological, connectivity, and shape indices

combination of  $A$  and  $m$ , there was a maximum and a minimum value of  ${}^mP$  that could be expressed in terms of  $A$ . Therefore, each molecule  $i$  has a shape and a series of related  ${}^mP_i$  values, each of which has a value between the two extremes of shape:

$${}^mP_{max} \geq {}^mP_i \geq {}^mP_{min}$$

where  $A$  is the same for all 3 structures.

Molecule	$A$	${}^1P$	
	6	15	Maximum possible value of ${}^1P$ for 6 atoms
	6	6	${}^1P_i$
	6	5	Minimum possible value of ${}^1P$ for 6 atoms

**Kappa Indices ( $\kappa$ )**

Each Kappa index is based on a mathematical expression of the ratios between  ${}^mP_{max}$ ,  ${}^mP_i$  and  ${}^mP_{min}$ .

Kappa 1 of  $\kappa$

Kappa 1 shows the degree of complexity of a bonding pattern.

For any number of atoms  $A$ :

$${}^1P_{max} = (A(A-1))/2 \quad \text{and} \quad {}^1P_{min} = (A-1)$$

Therefore, if

$${}^1\kappa = \frac{2 {}^1P_{max} {}^1P_{min}}{({}^1P_i)^2}$$

Then, substituting the equations for  ${}^1P_{max}$  and  ${}^1P_{min}$  into the equation:

Chapter 1 - Property Calculations

$${}^1\kappa = \frac{A(A-1)^2}{({}^1P_1)^2}$$

${}^1\kappa = A$  for an acyclic structure

${}^1\kappa > A$  for cyclic structures.

Kappa 2 or  ${}^2\kappa$  Kappa 2 indicates the degree of linearity or star-likeness of bonding patterns.

In a similar way, Kappa 2 may be defined as follows, for any number of atoms A:

$${}^2\kappa = \frac{(A-1)(A-2)^2}{({}^2P_1)^2}$$

Kappa 3 or  ${}^3\kappa$  Kappa 3 indicates the degree of branching at the centre of a molecule, larger for predominantly linear molecules with branching at the ends.

Kappa 3 may be defined as follows, for any number of atoms A:

$${}^3\kappa = \frac{(A-1)(A-3)^2}{({}^3P_1)^2} \quad \text{when A is odd}$$

$${}^3\kappa = \frac{(A-2)^2(A-3)}{({}^3P_1)^2} \quad \text{when A is even}$$



*Topological, connectivity, and shape indices***Kappa Alpha Indices ( $K_\alpha$ )**

The basic Kappa indices were derived assuming that all atoms are equivalent. Hall and Kier also defined a group of modified indices, the Kappa Alpha indices.

The contribution of each atom to the overall shape of a molecule is based on a comparison with a Carbon  $sp^3$  atom. An alpha value is calculated for each atom type using the ratio of Covalent radii of  $C(sp^3)$  and the atom.

$$\alpha_x = ((r_x / r_{C(sp^3)}) - 1)$$

For a molecule with non- $C(sp^3)$  atoms, the Kappa indices are modified by replacing  $A$  with  $(A + \alpha)$ , where  $\alpha$  is the sum of  $\alpha_x$  for all  $A$ . Therefore the Kappa Alpha equations are:

$${}^1K_\alpha = ((A + \alpha)(A + \alpha - 1)^2) / ({}^1P_i + \alpha)^2$$

$${}^2K_\alpha = ((A + \alpha - 1)(A + \alpha - 2)^2) / ({}^2P_i + \alpha)^2$$

$${}^3K_\alpha = ((A + \alpha - 1)(A + \alpha - 3)^2) / ({}^3P_i + \alpha)^2 \text{ when } A \text{ is odd}$$

$${}^3K_\alpha = ((A + \alpha - 2)^2(A + \alpha - 3)) / ({}^3P_i + \alpha)^2 \text{ when } A \text{ is even}$$

**Kappa Index values for small molecules**

In small molecules ( $A \leq 4$ ) some of the  ${}^mP$  quantities are not be defined, or are set to zero. Consequently, useful Kappa values cannot be calculated from the equations above.

Chapter 1 - Property Calculations

Hall and Kier extrapolated the expected indices for small molecules, using methods described in the 1992 reference:

$^1\kappa = 1.000$	for methane	$^3\kappa = 0.000$	for methane
$^2\kappa = 0.000$	for methane	$= 1.450$	for ethane
$= 1.000$	for ethane	$= 2.000$	for propane
		$= 3.378$	for butane

Hall and Kier do not discuss values for Kappa Alpha in small molecules. In Tsar, the Kappa Alpha values of small molecules are set to zero, unless  $\alpha=0.0$  (i.e., only  $C(sp^3)$  atoms), when they are set equal to the Kappa indices.

Flexibility or  $\phi$ 

Flexibility of a molecule is directly related to the degree of linearity, and the presence of cycles and/or branching. The Kappa alpha indices measure these factors while also taking the effects of atomic identities on shape into account.

Hall and Kier found that by combining  $^1\kappa_\alpha$  and  $^2\kappa_\alpha$  indices, a further index  $\phi$ , which measured flexibility, could be defined:

$$\phi = (^1\kappa_\alpha + ^2\kappa_\alpha) / A$$

For alkanes,  $\phi$  is equal to the number of bonds in the carbon skeleton. Branching and/or cyclisation lead to a sharp decline in  $\phi$ .